CSD-YOLOv8s：基于无人机图像的密集小目标羊只检测模型

翁　智[1,2]，刘海鑫[1,2]，郑志强[1,2*]

（1. 内蒙古大学 电子信息工程学院，内蒙古呼和浩特 010021，中国；2. 内蒙古大学草原家畜生殖调控与繁育国家重点实验室，内蒙古呼和浩特 010030，中国）

摘　要：［**目的/意义**］天然牧场下放牧性畜数量的准确检测是规模化养殖场改造升级的关键。为满足规模化养殖场对大批羊群实现精准实时的检测需求，提出一种高精度、易部署的小目标检测模型CSD-YOLOv8s（CBAM SPPFCSPC DSConv-YOLOv8s），实现无人机高空视角下小目标羊只个体的实时检测。［**方法**］首先，使用无人机获取天然草原牧场中包含不同背景及光照条件下的羊群视频数据并与下载的部分公开数据集共同构成原始图像数据。通过数据清洗和标注整理生成羊群检测数据集。其次，为解决羊群密集和相互遮挡造成的羊只检测困难问题，基于YOLO（You Only Look Once）v8模型构建具有跨阶段局部连接的SPPFCSPC（Spatial Pyramid Pooling Fast-CSPC）模块，提升网络特征提取和特征融合能力，增强模型对小目标羊只的检测性能。在模型的Neck部分引入了卷积注意力模块（Convolutional Block Attention Module, CBAM），从通道和空间两个维度增强网络的抗干扰能力，提升网络对复杂背景的抑制能力，进一步提高对密集羊群的检测性能。最后，为提升模型的实时性和可部署性，将Neck网络的标准卷积改为具有可变化内核的轻量卷积C2f_DS（C2f-DSConv）模块，减小了模型的参数量并提升了模型的检测速度。［**结果和讨论**］与YOLO、Faster R-CNN（Faster Regions with Convolutional Neural Networks）及其他经典网络模型相比，改进后的CSD-YOLOv8s模型在检测速度和模型大小相当的情况下，在羊群检测任务中具有更高的检测精度。Precision达到95.2%，mAP达到93.1%，FPS（Frames Per Second）达到87 f/s，并对不同遮挡程度的羊只目标具有较强的鲁棒性，有效解决了无人机检测任务中因羊只目标小、背景噪声大、密集程度高导致羊群漏检和误检严重的问题。公开数据集验证结果表明，提出的模型对其他不同物体的检测精度均有所提高，特别是在羊只检测方面，检测精度提升了9.7%。［**结论**］提出的CSD-YOLOv8s在无人机图像中更精准地检测草原放牧牲畜，对不同程度的聚集和遮挡目标实现精准检测，且具有较好的实时性，为养殖场大规模畜禽检测提供了技术支撑，具有广泛的应用潜力。

关键词：羊只检测；YOLOv8；小目标；SPPFCSPC；注意力机制；深度可分离卷积

中图分类号：TP391.41；S826　　**文献标志码：**A　　**文章编号：**SA202401004

引用格式：翁智，刘海鑫，郑志强. CSD-YOLOv8s:基于无人机图像的密集小目标羊只检测模型[J]. 智慧农业(中英文), 2024, 6(4): 42-52. DOI：10.12133/j.smartag.SA202401004

WENG Zhi, LIU Haixin, ZHENG Zhiqiang. CSD-YOLOv8s: Dense Sheep Small Target Detection Model Based on UAV Images[J]. Smart Agriculture, 2024, 6(4): 42-52. DOI：10.12133/j.smartag.SA202401004 (in Chinese with English abstract)

0　引　言

在草原牧场的日常管理中，牧群的便捷准确检测是牧场养殖者做出科学决策[1]、提高牧场收益的重要保障。早期阶段，牲畜调查采用有人驾驶飞机方式。这种方式有效地减少了时间投入，但由于高昂的费用、需要合格的飞行员及相关飞机设备等限制，应用于牧场受到一定制约。与有人驾驶的直升机相比，无人机具有更大的灵活性。牧场主只需一台无人机，就能每天监测牧场的牲畜，既能降低成本，又减轻了工作负担，为牧场管理提供了便捷经济的解决方案。如今，无人机已广泛应用于草原牧场放牧管理中。例如，赵建敏等[2]使用无人机对天然牧场中的羊群数量进行监测。在羊群密集场景

收稿日期：2024-01-05

基金项目：国家自然科学基金项目（61966026）；内蒙古自治区高等学校青年科技英才支持计划（NJYT23063）

作者简介：翁　智，研究方向为人工智能与机器视觉。E-mail： wzhi@imu.edu.cn

*通信作者：郑志强，博士，副教授，研究方向为图像处理。E-mail： zqzheng@imu.edu.cn

下实现了很好的羊群数量估计。但应用于复杂背景下的草原牧场羊群检测时，由于数据集场景较为单一、目标尺度变化不大，因此难以解决羊群的漏检和误检问题。Li等[3]将无人机作为“牧羊犬”，用于放牧牲畜。Herlin等[4]则将无人机与虚拟电子围栏结合实现高效放牧。无人机捕捉的图像与机器学习相结合，能够避免传统地面监测过程中环境及人为因素的影响[5]，可以为牧场主提供更加准确和全面的信息。然而，无人机羊群检测会受到背景复杂、光照条件多变及小目标等因素的影响。针对上述问题，研究有关算法模型是无人机羊群检测的关键。

近年来，深度学习在小目标检测领域得到广泛应用。郭秀明等[6]基于拆分不同尺度的数据图像，用迁移学习生成模型的方法提高小目标检测能力。Maktab等[7]采用反卷积和浅层特征融合的方法提升无人机小目标检测的精度。上述研究在小目标检测方面具有重要意义，提高了检测准确度。然而，考虑到无人机上的计算资源有限，因此提升模型检测精度并能实时检测是无人机对地小目标检测的关键点之一。此外，针对羊群目标检测，Zhao等[8]提出两阶段的实例分割模型，提高密集羊群的检测精度。然而在无人机羊群检测中，无人机飞行高度较高，羊群所占像素少，难以提取有用和易区分的特征。Sarwar等[9]使用基于区域卷积神经网络模型对无人机视角下的绵羊进行检测，在自制的数据集上通过实验设置最佳的训练参数，检测精度达到95.6%。Wang等[10]构建了增强的CSPDarknet（Cross Stage Partial Darknet）和加权聚合特征重提取金字塔模块，提高了无人机图像中羊群的检测性能。上述方法都提高了羊群检测精度，然而当羊群处于复杂背景下并发生聚集和遮挡时，小目标的羊只检测变得困难。

在对无人机图像中的牧场羊群进行检测时，与地面图像检测存在明显差异，主要有以下几个难点。首先，由于无人机飞行视野角度较大，羊只目标仅占有很少的像素，难以提取有用和易区分的特征进行检测。其次，无人机图像中光照条件多变且包含大量无效的复杂背景，存在如岩石、干草堆以及树木等干扰物体。最后，由于羊只姿势变化以及运动状态的变化，在无人机图像中产生不同的外观形态进一步增加了检测难度。为了解决上述问题，本研究设计了一种无人机视角下羊群检测模型CSD-YOLOv8s（CBAM SPPFCSPC DSConv-YOLOv8s）。主要贡献如下。

1）针对天然草原牧场下，无人机对地羊群检测任务中所面临羊只目标小、干扰物体多、羊群聚集和遮挡现象导致漏检和误检严重的问题，构建一种跨阶段特征连接的SPPFCSPC（Spatial Pyramid Pooling Fast-CSPC）结构，该结构将原始特征与快速空间金字塔池化网络输出特征相结合，充分保留模型的不同阶段的特征信息，有效地解决了羊群目标较小且遮挡严重问题，提升模型对羊群小目标的检测性能。

2）为解决羊只检测任务中复杂背景区域较大及羊群密集导致的羊只误检测问题，在特征提取结构上增加基于空间和通道两方面增强特征信息捕获的注意力机制，在空间上抑制背景信息，在通道上聚焦羊只目标，提高模型在复杂背景和不同光照条件下对多尺度羊群的检测能力，使得模型在复杂环境下对羊群的检测能力得到进一步的提升。

3）为平衡模型的检测精度和检测速度，以C2f模块为基础，将具有可变化内核的深度可分离卷积引入到模型中，区别于原模型的C2f卷积和传统的深度可分离卷积，提出的模块能够根据输入特征自适应地选择相应的卷积核进行特征提取，在更灵活地解决羊群检测过程中输入尺度变化问题的同时，减少模型的参数量和计算量，提高模型的推理速度，为将来模型的应用提供支持。

1　材料与方法

1.1 小目标羊群数据集构建

1.1.1　数据来源

为增加样本多样性、提高算法对不同环境的适应能力，本研究数据包括两部分。一部分为2023年7月在内蒙古赤峰市圣泉生态牧业有限公司牧场采集数据，选取天然牧场下的羊群作为拍摄对象。使用大疆MINI2无人机进行视频录制。将录制的视频通过Python的Imageio库进行分帧处理，每个视频获得300～900张独立的图像，共计15 876张。另一部分数据为RoboFlow网站上获取的小目标羊只公开数据，共计7 328张图像。两种数据的融合形成本研究数据集，提升模型的泛化能力。

为确保所采集数据的图像质量，在使用无人机进行羊群图像数据采集时，需要综合考虑多个关键因素以确保数据的有效性。首先，为减少光照对数据的影响，本研究在阴天及晴天等不同状况的天气下，分别在中午和傍晚等不同时段内进行羊群视频

数据的采集。其次，考虑到背景的多样性，本研究在牧场不同的放牧区域和背景下进行视频录制，确保数据的多样性和代表性。本研究在上述时间和场景下以25～70 m的高度录制视频。总计录制33个视频，每段视频时长为60～180 s，录制的视频格式为MP4（Moving Picture Experts Group 4），分辨率像素为2 560×1 440，帧率设置为30 f/s。在不同高度视频的示例帧如图1所示。



a. 示例帧高度为35 m



b. 示例帧高度为45 m



c. 示例帧高度为55 m



d. 示例帧高度为65 m

图1 无人机不同高度采集的羊群视频示例帧图像

Fig. 1 Example frame images of sheep video captured by UAVs at different altitudes

1.1.2 数据处理

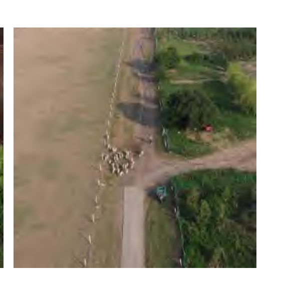
为确保数据图像的清晰度和多样性，删除了模糊和相似度较高的图像，方法如下。首先，离散化拉普拉斯算子（Laplace Operator）的二阶偏导数，并通过差分近似得到离散算子。叠加这些算子形成拉普拉斯矩阵（Laplacian Matrix），即卷积核。其次，用该卷积核依次计算所有图像方差，将计算结果与所设阈值（70）进行比较，删除低于阈值的图像。在处理图像相似度时，采用神经网络提取图像特征，并利用余弦相似度比较图像的相似度。通过与设定阈值（0.80）进行比较，删除相似度高于阈值的图像。通过上述处理得到3 862张图像，然后手动剔除羊只部分截断于图像边缘的数据，最终得到3 690张高质量的图像，按照编号0001～3690进行命名。数据分布情况见表1，数据集部分数据如图2所示。

表1 羊群小目标检测实验数据集分布

Table 1 Distribution of experimental datasets for sheep small target detection

数据集类型	图像数量/张	无人机采集数据/张	公开数据/张
训练集	2 214	1 584	630
验证集	738	528	210
测试集	738	528	210
总计	3 690	2 640	1 050





a. 无人机采集数据





b. 公开数据

图2 羊群数据集部分数据示例

Fig. 2 Example of data from the sheep dataset

使用LabelImg标注工具对图像进行标注，生成包含图像名称、宽高及目标位置等信息的XML标签文件。数据标注样例如图3所示。

1.2 基于CSD-YOLOv8s的密集羊只检测模型

YOLOv8模型在训练推理、检测精度和模型部



a. 原始标注样例图　　　　b. 标注样例展示图

图 3　羊群目标检测数据标注样例

Fig. 3 Sample data annotation for sheep target detection

署等方面有显著优势，因此逐渐应用于各种检测任务。在进行无人机羊只检测任务时，由于羊只目标尺寸小、易遮挡和聚集等问题，导致漏检测和误检测严重。因此，加强模型对小目标羊群的特征学习能力至关重要。为解决上述问题，通过改进 YOLOv8s 模型增强对无人机羊群检测的小目标检测能力，并在发生遮挡和聚集时提高羊只的检测精度，减少羊只误检测和漏检测数量。

首先在网络的主干部分构建基于原始快速空间金字塔池化网络的跨阶段局部网络结构（Cross Stage Partial, CSP）提升模型对小目标的特征提取能力。其次，引入了卷积注意力模块，在模型的第 17、21 和 25 层之后分别添加了卷积注意力模块（Convolutional Block Attention Module, CBAM）[11]，提高网络对不同通道和空间位置的关注能力，使网络更加关注羊只区域，从而改善网络的性能和泛化能力，进一步提高模型对密集羊只的检测性能。最后，为增加模型的检测速度，提升模型的可部署性，构建具有可变化内核的 C2f_DS 模块替换 YOLOv8 的 C2f 卷积模块，该模块具有较少的参数和更低的计算复杂度，能够在一定程度上提高模型的训练速度。改进后的模型命名为 CSD-YOLOv8s（CSD 分别是 CBAM、SPPFCSPC 和 DSConv 的首字母），其结构如图 4 所示。

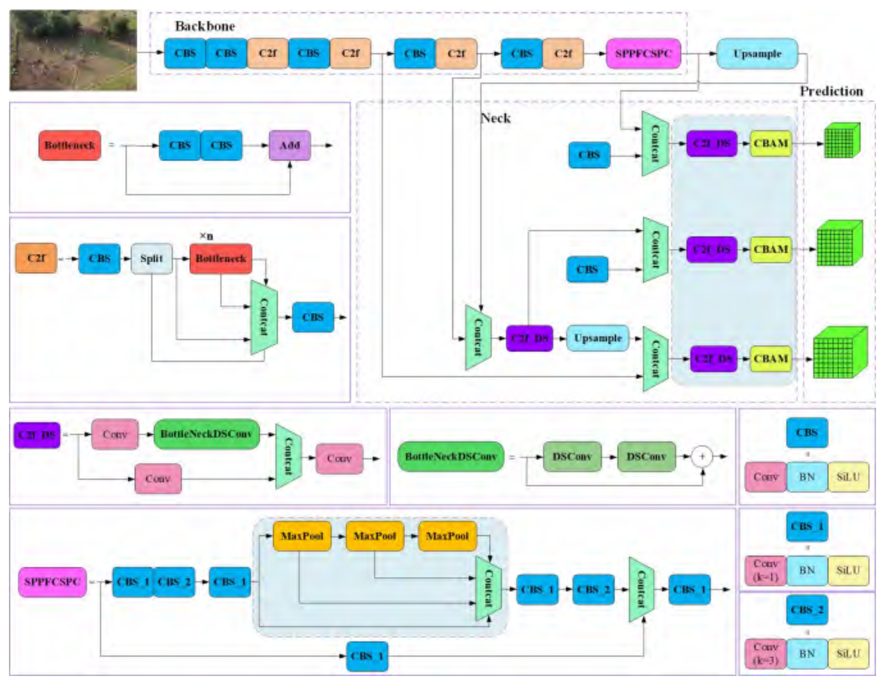


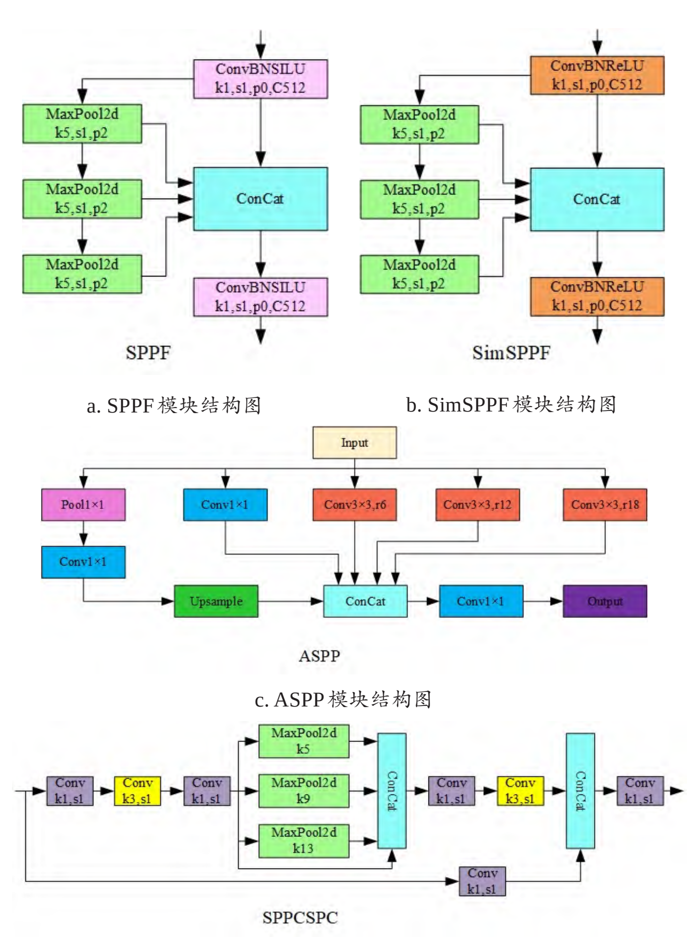
图 4　CSD-YOLOv8s 羊群小目标检测模型结构

Fig. 4 Model structure of the proposed CSD-YOLOv8s for sheep small target detection

1.2.1　SPPFCSPC 增强特征提取

在羊只检测任务中，由于背景、光线、尺度变化及遮挡聚集等因素，对小目标的检测能力和检测速度都提出更高的要求。随着网络层数的加深，羊只小目标的特征表示逐渐减弱，造成漏检测和误检测。在现阶段研究中，通常采用空洞金字塔池化（Atrous Spatial Pyramid Pooling, ASPP）、SimSPPF（Simplified Spatial Pyramid Pooling Fast）、SPPCSPC（Spatial Pyramid Pooling-CSPC）等方法对 YOLOv8 中的 SPPF（Spatial Pyramid Pooling Fast）模块进行改进，上述四种网络结构如图 5 所示。ASPP 和 SimSPPF 两种改进策略保证了模型检测的实时性，应用 SPPCSPC 模块能增强模型的特征融合能力，提高模型检测精度。

上述方法均未能兼顾检测精度和检测速度的需求，在实际羊群检测任务中无法达到快速、精确的



a. SPPF 模块结构图 b. SimSPPF 模块结构图

c. ASPP 模块结构图

d. SPPCSPC 模块结构图

图5 不同SPPF模块改进方法结构图

Fig. 5 Structure diagrams of different SPPF module improvement methods

检测。为解决上述问题，本研究在网络的主干部分采用基于原始快速空间金字塔池化的跨阶段局部连接网络结构。通过将CSP引入到SPPF中增强网络的特征提取能力，提高小目标及目标遮挡情景下模型精准检测的能力。改进的SPPFCSPC结构如图6所示。SPPFCSPC模块将输入特征进行拆分，部分特征进行SPPF结构处理，通过在多个卷积操作过程中使用串行的池化操作增强模型对不同尺度的处理能力，其余特征则直接经过卷积操作后与SPPF模块的输出相结合。这种结构将输入特征与输出特征进行深度结合，全面整合网络的深浅层特征并优化了梯度变化，牺牲较小的速度提高模型的稳定性和准确性。

1.2.2 注意力机制

当前构建模型的主要注意力机制有挤压注意力模块[12]（Squeeze-and-Excitation, SE）、渠道注意力模块[13]（Coordinate Attention, CA）及高效通道注意力模块[14]（Efficient Channel Attention, ECA）等。在羊群检测任务中，聚集的羊群和复杂的背景导致羊群检测过程中有效信息提取困难，影响模型的羊只检测结果，为解决复杂背景和不同光照条件下羊群聚集导致漏检严重的问题，本研究在网络的特征提取层加入了基于空间和通道两方面增强特征信息捕获的CBAM注意力机制。CBAM注意力模块由通道注意力模块（Channel Attention Module, CAM）和空间注意力模块（Spatial Attention Module, SAM）构成，从空间和通道两方面提升目标的关键特征提取能力。在无人机羊群检测时，羊群周围存在大量干扰检测的背景信息，CBAM通过空间注意力抑制背景干扰，使网络更专注于羊只目标。由于羊只属于小目标，特征信息少且易聚集遮挡，CBAM通过学习特征图的通道和空间注意力，自适应地增强重要的特征表示，更加关注小目标的特征，提升网络对局部和全局信息的感知能力。从而更好地检测羊只目标，提高整体检测准确率。CBAM模块整体流程如图7所示。

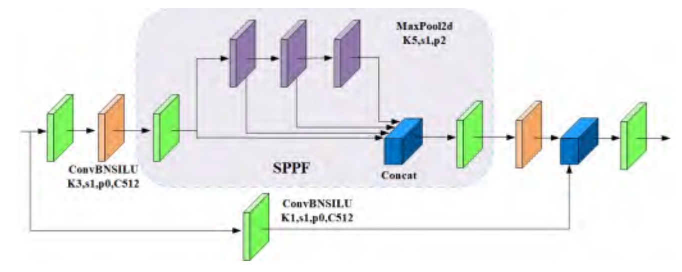


图6 改进的SPPFCSPC结构图

Fig. 6 Structure diagram of the improved SPPFCSPC

1.2.3 深度可分离卷积的模型轻量化

深度可分离卷积（Depthwise Separable Convolution, DSConv）由深度卷积（Depthwise Convolution, DW）和逐点卷积（Pointwise Convolution, PW）两部分组成[15]。通过引入可变化内核优化了DSConv模块以提高运行速度，同时增加模型的灵活性以适应复杂的数据。改进后的C2f_DS模块能够根据输入特征自适应地选择最合适的卷积核进行特征提取。改进后的C2f_DS模块包括深度卷积和逐点卷积两部分，前者通过独立卷积核对每个通道卷积，并将输出进行堆叠以减少参数，后者则用相应的卷积核对深度卷积输出进行通道混合生成最终特征图。

在羊只个体检测中，需要较小的模型及较高的推理速度以满足实时性和易部署的要求。因此本研究在YOLOv8s中构建C2f_DS模块代替标准卷积，

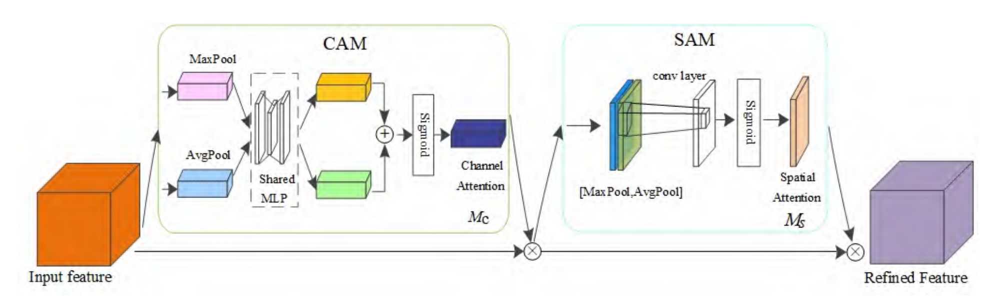
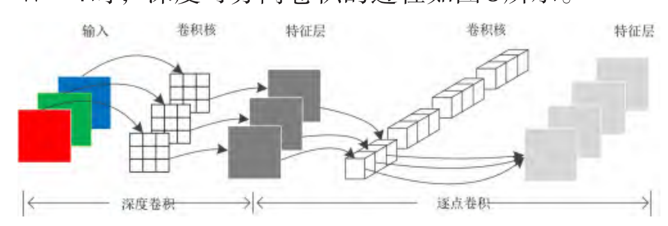


图7　CBAM模块总体流程图

Fig. 7 General flowchart of CBAM module

降低模型大小和计算量，提高模型的检测速度。当输入特征维度为H×W×3，输出特征维度为H×W×4时，深度可分离卷积的过程如图8所示。



注：输入特征维度为H×W×3，输出特征维度为H×W×4。

图8　深度可分离卷积过程图

Fig. 8 DSConv process diagram

1.3 评价指标

为了描述模型的检测效果，采用精度（Precision, P）、全类平均精度（mean Average Precision, mAP）、平均帧率（Frames Per Second, FPS）、参数量（Params）以及计算量（FLOPs）作为目标检测算法的评价指标。FPS表示模型1 s内处理的图像数量。Params用于衡量模型的复杂度。FLOPs用于衡量模型的计算量。AP值基于精度召回（Recall, R）曲线，计算该曲线下的面积并求均值。mAP考虑了检测算法的召回率和准确率，将每类的AP值进行平均。AP和mAP计算如公式（1）和公式（2）所示；P、R计算如公式（3）和公式（4）所示。在羊群检测任务中，四个核心评价指标为：TP（成功预测为羊的数量）、TN（成功预测为非羊的数量）、FP（错误预测为羊的数量）、FN（错误预测非羊的数量）。

$$\mathrm{AP}=\int_{0}^{1}P(R)dR \tag{1}$$

$$\mathrm{mAP}=\frac{1}{c}\sum_{j=1}^{c}\mathrm{AP}_{j} \tag{2}$$

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\times 100\% \tag{3}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\times 100\% \tag{4}$$

本研究中指定评价指标的优先级顺序为P、mAP、FPS、FLOPs、Params，以便于对模型进行评估。

2　结果分析

2.1 实验环境及参数设置

本实验所用的硬件设备及参数如表2所示。在模型训练阶段，使用2 214张图像作为训练集，使用738张图像作为验证集来评估模型性能并进行调整，从而获得最佳的羊只目标检测模型。为适应模型的输入，将原始图像像素调整为640×640、批量大小（Batch Size）设置为32，选择SGD（Stochastic Gradient Descent）作为优化器，动量（Momentum）设定为0.937，初始学习率为0.001。在上述硬件和参数设置下，进行了200个epochs的模型训练。

表2　无人机羊群目标检测实验硬件参数配置

Table 2 Hardware parameter configuration for UAV sheep target detection experiment

硬件配置	参数
CPU	Intel（R）Core（TM）i9-9900K CPU@3.6 GHz
GPU	NVIDIA Quadro P6000
内存容量	32 G
操作系统	Windows 10
开发工具	PyCharm2020.1.1
CUDA版本	10.2
深度学习框架	PyTorch 1.10.0

2.2 消融测试性能

为验证各部分改进方法的有效性，本研究设计消融实验验证SPPFCSPC模块、CBAM及C2f_DS模块对整体网络的影响。在同一数据集和软硬件设备下，将不同模块分别加入YOLOv8s模型中，评估各模块对网络性能的影响，评估结果如表3所示。

表3 无人机羊群目标检测消融实验

Table 3 UAV sheep target detectionAblation experiment

模型名称	*P*/%	mAP/%	FLOPs/G	FPS/(f/s)	Params/M
YOLOv8s	93.0	91.2	28.6	105	11.13
YOLOv8s+SPPFCSPC	94.6	92.6	33.6	93	17.49
YOLOv8s+CBAM	94.1	92.1	28.7	88	11.17
YOLOv8s+C2f_DS	92.8	91.2	24.7	107	10.31
YOLOv8s+SPPFCSPC+CBAM	95.2	93.1	33.9	82	17.61
YOLOv8s+SPPFCSPC+C2f_DS	94.5	92.4	29.9	95	16.73
YOLOv8s+CBAM+C2f_DS	94.2	92..0	24.8	90	10.36
CSD-YOLOv8s(本研究)	95.2	93.1	29.9	87	16.68

从表3中可知，在YOLOv8s模型上将SPPFCSPC替换原模型的SPPF模块后，模型的*P*值和mAP值分别提高了1.6和1.4个百分点，可见采用跨阶段特征连接结构能够充分保留不同阶段的特征信息，大幅度提升了模型的精度。添加CBAM注意力机制后与原始模型相比，*P*和mAP分别提高了1.1和0.9个百分点，体现了CBAM注意力机制在空间上对无用的背景信息的抑制，在通道上提高对羊群小目标的聚焦能力，与此同时，FPS减少了17 f/s，参数量增加了0.04 M，FLOPs增加了0.1 G。使用C2f_DS模块与原模型相比，*P*值降低了0.2%，但FPS提高了2 f/s且FLOPs减少了3.9 G，有效地证明了本研究提出的具有可变化内核的C2f_DS模块能够有效提取特征信息的同时降低了模型的计算量和参数量，提高了模型的检测速度。在加入SPPFCSPC与CBAM注意力机制后，模型的检测精度提升了2.2个百分点，参数量和计算量分别提升了5.3 G和6.48 M，与此同时，FPS下降了23 f/s，在加入C2f_DS模块后，模型的精度值保持不变，但降低了计算量和参数量增加了模型的推理速度。在C2f_DS模块分别加入SPPFCSPC和CBAM注意力机制后，对模型的精度影响不大，但降低了模型的参数量和计算量，并提高了模型的检测速度，证明了C2f_DS模块在提高计算效率上的有效性。本研究提出的CSD-YOLOv8s模型在无人机对地羊群目标检测任务中P值达到95.2%，mAP达到93.1%。与原始的YOLOv8s模型相比，*P*和mAP分别提高了2.2和1.9个百分点，同时FPS为87 f/s，Params为16.68 M，FLOPs为29.9 G，能够在后期将其部署在无人机的机载电脑上，为实现目标的实时检测提供参考。

2.3 对比实验结果分析

2.3.1 不同模型对比

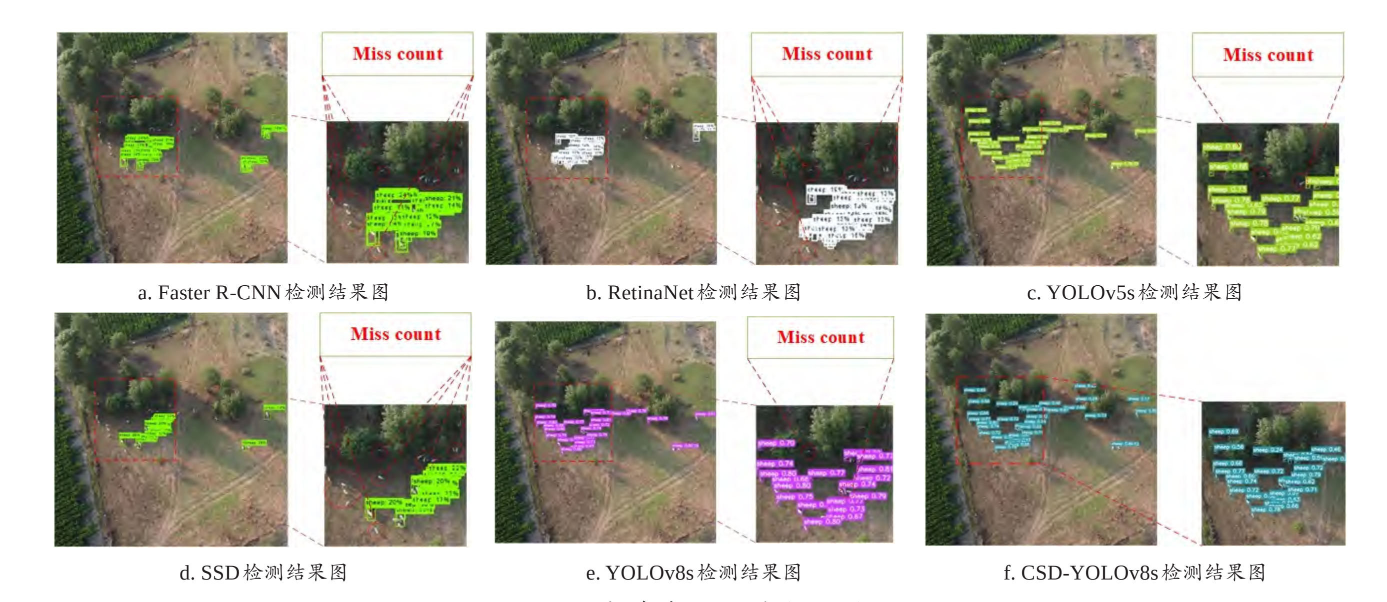
选取不同模型，验证和比较所提出的方法对无人机对地小目标检测的性能。选取单阶段目标检测算法SSD[16]（Single Shot MultiBox Detector）和RetinaNet[17]模型，双阶段Faster R-CNN[18]模型，YOLOv5s以及YOLOv8s模型与CSD-YOLOv8s模型进行性能对比实验，结果如图9所示。

不同模型在测试集上的mAP、FPS和Model size结果如表4所示。本研究提出模型CSD-YOLOv8s与Faster R-CNN模型相比，mAP提高了28.5个百分点，平均帧率增加了69 f/s，模型大小减少了279 M；与RetinaNet、SSD模型相比，mAP提高了16.9、34.3个百分点，FPS提高了69f/s、28 f/s；模型大小分别减少了210 M和65 M。与YOLOv5s以及YOLOv8s模型相比，FPS有所降低，Model size有所增长，但mAP提高了2.6和1.9个百分点。由表中可知，CSD-YOLOv8s能够大幅度地提高检测精度的同时保证模型大小和检测速度的均衡，检测效果优于SSD、REtinaNet、经典的Faster R-CNN以及比较流行的YOLO系列算法，可为高空实时检测羊群提供参考。

2.3.2 不同注意力机制对比

注意力机制能使模型更专注感兴趣的目标，提高权重聚焦重要信息。通过对比不同注意力机制，匹配羊群检测任务中最佳的注意力机制设计方案，表5对几种常用注意力机制实验结果进行了统计，重点关注其准确率等参数。

本研究通过引入CAM和SAM来提高模型提取特征的能力，以适应复杂的环境。实验结果表明，CBAM注意力机制对于提升模型的准确率最为明显，与未使用注意力机制相比，参数量有所增长，



a. Faster R-CNN检测结果图　b. RetinaNet检测结果图　c. YOLOv5s检测结果图

d. SSD检测结果图　e. YOLOv8s检测结果图　f. CSD-YOLOv8s检测结果图

图9　无人机羊群检测同模型检测结果

Fig. 9　UAV flock detection with modeling results

表4　不同网络模型的羊群检测实验结果

Table 4　Experimental results of sheep detection with different network models

Model	mAP/%	FPS/(f/s)	Model size/M
Faster R-CNN	64.6	18	314
REtinaNet	76.2	18	245
YOLOv5s	90.5	115	14
SSD	58.8	59	100
YOLOv8s	91.2	105	22
CSD-YOLOv8s	93.1	87	35

表5　不同注意力机制实验对比结果

Table 5　Experimental comparison of different attention mechanisms

Attention mechanism	*P*/%	mAP/%	FLOPs/G	FPS/(f/s)
——	93.0	91.2	28.6	105
SE	93.7	91.5	28.7	96
SimAM	93.2	91.2	28.6	100
CA	93.2	91.4	28.7	96
Shuffle	92.8	91.1	28.6	116
ECA	93.3	91.2	28.7	101
CBAM	94.1	92.1	28.7	88

注："——"代表未加注意力机制的实验结果。

但是P值提高了1.1个百分点，mAP增加了0.9个百分点。与其他注意力机制相比，CBAM注意力机制在结构上增加了空间注意力模块，虽然参数量有所增长，但是显著增强了模型从空间和通道两方面提取羊只特征的能力，在一定程度上提高了检测精度。与SE相比，CBAM的CAM结构通过增加一个并行的最大池化层，得到更为丰富的特征信息，因此在复杂背景及光照条件下，增强了对羊只的检测精度，并减少了误检测。在羊群检测任务中，全局信息对于正确定位和检测羊只至关重要，而SimAM注意力机制过于关注特征图中不同位置之间的相似度，忽略了全局信息，因此影响了对羊只目标的检测性能。CA注意力机制和ECA注意力机制则主要关注不同通道之间的关系而忽略了空间信息，导致网络对小目标羊只在图像中的不同位置和尺度的适应能力不足。Shuffle[19]注意力机制基于通道随机重排。在羊群检测任务中，即使很少的通道也会包含大量的羊只信息，而随机重排使得这些通道被打乱或者忽略，从而影响网络对于羊只的检测能力。

2.3.3　改进SPPF模块的试验分析

在YOLOv8中，SPPF模块是一种关键的特征提取结构，其主要目的是增强目标检测模型对不同尺度目标的感知能力。本研究为验证改进后的SPPFCSPC特征提取模块对无人机羊群检测任务中的高效性，对现有的SPPF改进方法进行了对比验证，实验结果如表6所示。

本研究将CSP跨阶段局部网络连接引入到SPPF模块中，与原模型相比，*P*值和mAP分别提高了1.6和1.4个百分点。改进后的SPPFCSPC模块将原始特征与SPPF模块输出特征相结合，保留了更多的特征信息，虽然增加了模型的参数量，但提高了小目标的特征提取能力，特别是在羊只密集和遮挡严重的情况下，增加了羊只检测的准确性，降低了羊只目标的漏检和误检率。ASPP模块是一种基

表6 不同SPPF模块实验对比结果

Table 6 Experimental pairwise comparisonresults of different SPPF modules

Module	*P*/%	mAP/%	FLOPs/G	FPS/(f/s)	Params/M
SPPF	93.0	91.2	28.6	105	11.13
ASPP	93.3	91.0	35.0	84	19.38
SimSPPF	93.0	91.2	28.6	106	11.14
SPPCSPC	94.6	92.6	33.6	90	17.56
SPPFCSPC	94.6	92.6	33.6	93	17.56

于较大的感受野用于处理不同尺度目标的卷积神经网络模块，在羊群检测中由于目标尺寸小，且羊群遮挡和聚集严重从而导致检测效果不佳。SimSPPF模块替换了SPPF模块的激活函数，增加了速度，但特征提取能力有限，在羊只检测任务中无法有效捕捉到羊只目标。SPPCSPC模块在YOLOv7中被提出[20]，提高了模型的检测性能，但使用多个并联结构的池化操作增加了模型的重复运算减小了模型的运算速度，因此存在计算效率较低、实时性不足的问题。与其他改进方法相比，本研究的改进思路充分考虑了网络的不同层次特征，通过融合深浅层特征提高了模型的学习能力。实验结果表明，使用SPPFCSPC方法能够提升精度值并将FPS和FLOPs保持在不错的水平以满足实时性的要求。

2.3.4 公开数据集验证

为了验证CSD-YOLOv8s的实用性和泛化能力，采用CSD-YOLOv8s模型与YOLOv8s模型在公共数据集PASCAL VOC 2007上进行对比实验（表7）。PASCAL VOC 2007数据集包含交通工具、人及动物等20种不同类别的图像数据并经过详细地标注，应用于计算机视觉领域进行算法评估。结果表明，采用CSD-YOLOv8s模型与YOLOv8s模型相比，所有类别（all）的*P*值提高了2.1个百分点，mAP增加了0.8个百分点。其中羊（sheep）的检测精度有大幅度的提升，*P*值提高了9.7个百分点，mAP提高了1.1个百分点。由此可见，本研究通过结合三种改进策略所提出的模型表现出较好的泛化能力，能够适用各种场景下的物体检测任务。尤其是在农业环境的羊只检测领域，该模型表现了卓越的性能，为羊只检测任务提供了一种高效的方法。

表7 公共数据集PASCAL VOC 2007对比实验结果

Table 7 The comparison experiment results of public dataset PASCAL VOC 2007

Model	*P*/%	mAP/%
YOLOv8s(all)	0.763	0.702
YOLOv8s(sheep)	0.722	0.652
CSD-YOLOv8s(all)	0.784	0.710
CSD-YOLOv8s(sheep)	0.819	0.663

注：all代表数据集中所有类别；sheep表示数据集中“羊”的类别。

3 讨论与结论

3.1 讨论

本研究基于YOLOv8s提出了一种高精度、实时性的CSD-YOLOv8s模型，解决了在复杂背景及光照条件下无人机对地羊群小目标检测过程中存在漏检和误检严重的问题。在羊群小目标检测过程中，发现在距离较远、羊群密集及遮挡情况下检测效果不佳，为此构建了跨阶段特征融合的网络结构，充分考虑模型深浅层特征，提升对遮挡目标的检测能力。融入注意力机制，关注模型的全局信息提升对小目标的检测能力，并使用轻量化卷积C2f_DS模块保证模型的检测速度。

CSD-YOLOv8s模型在羊群密集和遮挡情况下实现了很好的检测，但是与YOLOv8模型相比，FPS指标有所降低。这主要是跨阶段特征融合需要进行更多的卷积操作，增加了模型的计算量和参数，而融入注意力机制不仅增加了内存占用，还提高了模型的计算复杂度，因此模型推理需要更多的计算资源，增加了推理时间。在实际应用中，可以通过知识蒸馏等方法减少模型的大小和计算量从而在保证精度的同时缩短推理时间。此外，在将模型应用到如岩石较多的特殊环境时，可能会增加误检测。这是由于本研究的羊只数据主要集中在白色或灰色，和牧草旺盛时期的绿色背景颜色对比鲜明，从而在一定程度上提高了检测的准确性。这可以考虑增加特殊环境的数据或利用图像增强技术生成伪图像等方法解决。

除了对上述问题进行探索，未来仍有需改进的方面。首先，通过增加不同国家和地区牧场的羊群数据提高模型的检测能力，扩展模型的应用场景。其次，增加牧区里其他放牧种类如牛、马等以探究物体检测领域中复杂的长尾对象检测问题，扩展模型的应用范围。最后，为进一步提高应用，实现牧场的智能化放牧，将模型部署到无人机的机载电脑上，通过无人机实现牧场放牧牲畜的自动检测和跟踪，更好地将深度学习融入遥感，服务于畜牧业。

3.2 结论

本研究针对无人机对地目标检测中存在目标小，密集以及遮挡严重导致漏检和误检的问题，构

建了基于无人机图像的羊群数据集，提出一种用于天然草原放牧羊群小目标检测的深度学习模型CSD-YOLOv8s，获得较好的检测结果。首先，以YOLOv8s模型为基础，构建将原始特征与快速空间金字塔池化网络输出相结合的SPPFCSPC模块并融入CBAM注意力机制，增强模型的关键特征提取能力。其次，采用C2f_DS轻量化卷积模块缩减模型的参数量和计算量。最后在自建数据集上进行实验，结果表明，本研究提出的模型对羊群个体识别的mAP为93.1%，FPS为87 f/s，Params为16.68 M，FLOPs为29.9 G。

改进后的模型在无人机羊群检测任务中对密集及遮挡下的羊只检测精度有不错的效果提升，经PASCAL VOC 2007公开数据集验证，本研究提出的检测模型对包含动物、交通工具等20种类别物体的检测精度均有所提高，特别是在羊只检测方面，检测精度提升9.7个百分点。本研究提出的模型有效地解决了无人机检测任务中羊群漏检和误检现象严重的问题，为天然牧场无人机视角下羊群的检测提供了一种有效的检测方法。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] AQUILANI C, CONFESSORE A, BOZZI R, et al. Review: Precision livestock farming technologies in pasture-based livestock systems[J]. Animal, 2022, 16(1): ID 100429.

[2] 赵建敏, 李雪冬, 李宝山. 基于无人机图像的羊群密集计数算法研究[J]. 激光与光电子学进展, 2021, 58(22): ID 2210013.
ZHAO J M, LI X D, LI B S. Algorithm of sheep dense counting based on unmanned aerial vehicle images[J]. Laser & optoelectronics progress, 2021, 58(22): ID 2210013.

[3] LI X H, HUANG H L, SAVKIN A, et al. Robotic herding of farm animals using a network of barking aerial drones[J]. Drones, 2022, 6(2): ID 29.

[4] HERLIN A, BRUNBERG E, HULTGREN J, et al. Animal welfare implications of digital tools for monitoring and management of cattle and sheep on pasture[J]. Animals: an open access journal from MDPI, 2021, 11(3): ID 829.

[5] CORCORAN E, DENMAN S, HAMILTON G. Evaluating new technology for biodiversity monitoring: Are drone surveys biased?[J]. Ecology and evolution, 2021, 11(11): 6649-6656.

[6] 郭秀明, 诸叶平, 李世娟, 等. 农业复杂环境下尺度自适应小目标识别算法——以蜜蜂为研究对象[J]. 智慧农业(中英文), 2022, 4(1): 140-149.
GUO X M, ZHU Y P, LI S J, et al. Scale adaptive small objects detection method in complex agricultural environment: Taking bees as research object[J]. Smart agriculture, 2022, 4(1): 140-149.

[7] MAKTAB DAR OGHAZ M, RAZAAK M, REMAGNINO P. Enhanced single shot small object detector for aerial imagery using super-resolution, feature fusion and deconvolution[J]. Sensors, 2022, 22(12): ID 4339.

[8] ZHAO H K, MAO R, LI M, et al. SheepInst: A high-performance instance segmentation of sheep images based on deep learning[J]. Animals: An open access journal from MDPI, 2023, 13(8): ID 1338.

[9] SARWAR F, GRIFFIN A, PERIASAMY P, et al. Detecting and counting sheep with a convolutional neural network[C]// 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Piscataway, New Jersey, USA: IEEE, 2018: 1-6.

[10] WANG Y H, MA L L, WANG Q, et al. A lightweight and high-accuracy deep learning method for grassland grazing livestock detection using UAV imagery[J]. Remote sensing, 2023, 15(6): ID 1593.

[11] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]// European Conference on Computer Vision. Cham, German: Springer, 2018: 3-19.

[12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 7132-7141.

[13] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2021: 13708-13717.

[14] WANG Q L, WU B G, ZHU P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2020: 11531-11539.

[15] Depthwise separable convolutions for machine learning[EB/OL]. Eli Bendersky. (2018-04-04) [2023-12-28]. https://eli. thegreenplace. net/2018/depthwise-separable-convolutions-for-machine-learning/.

[16] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[M]// LEIBE B, MATAS J, SEBE N, et al, eds. Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

[17] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2017: 2999-3007.

[18] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137-1149.

[19] ZHANG Q L, YANG Y B. SA-net: Shuffle attention for deep convolutional neural networks[C]// ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, New Jersey, USA: IEEE, 2021: 2235-2239.

[20] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition

(CVPR). Piscataway, New Jersey, USA: IEEE, 2023: 7464-7475.

CSD-YOLOv8s: Dense Sheep Small Target Detection Model Based on UAV Images

WENG Zhi[1,2], LIU Haixin[1,2], ZHENG Zhiqiang[1,2*]

(1. *School of Electronic Information Engineering, Inner Mongolia University, Hohhot* 010021, *China*; 2. *State Key Laboratory of Grassland Livestock Reproduction Regulation and Breeding, Inner Mongolia University, Hohhot* 010030, *China*)

Abstract:

[Objective] The monitoring of livestock grazing in natural pastures is a key aspect of the transformation and upgrading of large-scale breeding farms. In order to meet the demand for large-scale farms to achieve accurate real-time detection of a large number of sheep, a high-precision and easy-to-deploy small-target detection model: CSD-YOLOv8s was proposed to realize the real-time detection of small-targeted individual sheep under the high-altitude view of the unmanned aerial vehicle (UAV).

[Methods] Firstly, a UAV was used to acquire video data of sheep in natural grassland pastures with different backgrounds and lighting conditions, and together with some public datasets downloaded formed the original image data. The sheep detection dataset was generated through data cleaning and labeling. Secondly, in order to solve the difficult problem of sheep detection caused by dense flocks and mutual occlusion, the SPPFCSPC module was constructed with cross-stage local connection based on the you only look once (YOLO)v8 model, which combined the original features with the output features of the fast spatial pyramid pooling network, fully retained the feature information at different stages of the model, and effectively solved the problem of small targets and serious occlusion of the sheep, and improved the detection performance of the model for small sheep targets. In the Neck part of the model, the convolutional block attention module (CBAM) convolutional attention module was introduced to enhance the feature information capture based on both spatial and channel aspects, suppressing the background information spatially and focusing on the sheep target in the channel, enhancing the network's anti-jamming ability from both channel and spatial dimensions, and improving the model's detection performance of multi-scale sheep under complex backgrounds and different illumination conditions. Finally, in order to improve the real-time and deploy ability of the model, the standard convolution of the Neck network was changed to a lightweight convolutional C2f_DS module with a changeable kernel, which was able to adaptively select the corresponding convolutional kernel for feature extraction according to the input features, and solved the problem of input scale change in the process of sheep detection in a more flexible way, and at the same time, the number of parameters of the model was reduced and the speed of the model was improved.

[Results and Discussions] The improved CSD-YOLOv8s model exhibited excellent performance in the sheep detection task. Compared with YOLO, Faster R-CNN and other classical network models, the improved CSD-YOLOv8s model had higher detection accuracy and frames per second (FPS) of 87 f/s in the flock detection task with comparable detection speed and model size. Compared with the YOLOv8s model, Precision was improved from 93.0% to 95.2%, mAP was improved from 91.2% to 93.1%, and it had strong robustness to sheep targets with different degree of occlusion and different scales, which effectively solved the serious problems of missed and misdetection of sheep in the grassland pasture UAV-on-ground sheep detection task due to the small sheep targets, large background noise, and high degree of densification. misdetection serious problems. Validated by the PASCAL VOC 2007 open dataset, the CSD-YOLOv8s model proposed in this study improved the detection accuracy of 20 different objects, including transportation vehicles, animals, etc., especially in sheep detection, the detection accuracy was improved by 9.7%.

[Conclusions] This study establishes a sheep dataset based on drone images and proposes a model called CSD-YOLOv8s for detecting grazing sheep in natural grasslands. The model addresses the serious issues of missed detections and false alarms in sheep detection under complex backgrounds and lighting conditions, enabling more accurate detection of grazing livestock in drone images. It achieves precise detection of targets with varying degrees of clustering and occlusion and possesses good real-time performance. This model provides an effective detection method for detecting sheep herds from the perspective of drones in natural pastures and offers technical support for large-scale livestock detection in breeding farms, with wide-ranging potential applications.

Key words: sheep detection; YOLOv8; small target; SPPFCSPC; attention mechanism; depthwise separable convolution

Foundation items: National Natural Science Foundation of China Under Grant (61966026); Program for Young Talents of Science and Technology in Universities of Inner Mongolia Autonomous Region Grant (NJYT23063)

Biography: WENG Zhi, E-mail: wzhi@imu.edu.cn

***Corresponding author:** ZHENG Zhiqiang, E-mail: zqzheng@imu.edu.cn